Sankhyā : The Indian Journal of Statistics
1992, Volume **54**, Series A, Pt. 3, pp. 387-401.

A CRAMER–RAO TYPE INEQUALITY FOR RANDOM VARIABLES IN EUCLIDEAN MANIFOLDS

By H. W. M. HENDRIKS, J. H. M. JANSSEN and F. H. RUYMGAART*
Katholieke Universiteit Nijmegen
and
Texas Tech University

SUMMARY. In this note the Cramér-Rao type inequality for estimators with values in an abstract manifold in Hendriks (1987) is specialized to manifolds in $\mathbf{R}^k$. Compared with the classical case the lower bound has a geometrical interpretation in terms of the Weingarten map. In passing some useful concepts regarding location and dispersion and the potential use of the exponential mapping are discussed. The examples focus on the special orthogonal group that plays a role in Chang (1986), with the circle (see Mardia, 1972) as a special case. The sphere (see Watson, 1983) is also considered.

1. Introduction : some basic concepts

In Hendriks (1987) a Cramér-Rao type lower bound has been established for estimators with values in an abstract manifold. Here we will give some more examples by first specializing to submanifolds of $\boldsymbol{R}^k$, $k \in \boldsymbol{N}$ referred to as Euclidean manifolds in the title. This restriction entails that the Euclidean metric and ordinary (componentwise) differentiation of (vector valued) functions can be used, so that an elementary formulation and proof of the inequality can be given ; see Section 2. On the other hand this special case is still general enough to cover most situations that are of practical importance. It is interesting to note that the lower bound contains a kind of correction term as compared with the standard situation, that has an interpretation in terms of the Weingarten map for arbitrary codimension. In Section 3 we will in particular pay attention to the special orthogonal group, that plays a role in Chang (1986) ; we give the inequality and briefly touch upon the attainment of the lower bound in this case. A special case is the circle (Mardia, 1972).

For the formulation of the inequality we need to briefly review some

Paper received. January 1988 ; revised October 1990.

AMS (1980) *subject classification.* Primary **62H12** ; secondary **22E99.**

Key words and phrases. Cramér-Rao inequality, manifold, special orthogonal group, circle, sphere.

*Part of this work was done while this author was visiting the Department of Statistics, UNC, Chapel Hill.

concepts in Hendriks (1987) in the present setting and use a somewhat different terminology. In subsections 1.2 and 1.3 we introduce the mean location and divergence respectively of a random variable with values in a manifold in $\boldsymbol{R}^k$. These concepts are of general importance and the mean location coincides with the expectation if the manifold is a convex set for instance. The mean location, moreover, specializes to the mean direction as defined for the circle and the sphere (Mardia, 1972, Watson, 1983). The divergence is related to Mardia's (1972) concept of divergence for the circle, as will be shown below. In subsection 1.1 we will briefly discuss the exponential mapping as a tool to construct examples of probability measures and to deal with weak convergence on certain classes of manifolds, like the special orthogonal group in particular.

1.1. *Exponential map and wrapping.* Given the tangent space at a certain point of the manifold, the exponential mapping maps a neighborhood of the origin in the tangent space onto a neighborhood of the point on the manifold, such that the mapping is one-to-one and maps straight lines through the origin on geodesics through that point preserving distance. Although in general the mapping is defined only locally, in some cases it is defined on the entire tangent space and maps onto the entire manifold, where the one-to-one and distance preserving character is usually lost in the global sense. In such cases probability measures on the linear tangent space are naturally transformed into probabilities on the manifold. Moreover, such linear operations on elements in the tangent space as often needed for asymptotic normality, result in an element that can be mapped in the manifold. In such a way weak convergence on the manifold might be defined. Here we will not be concerned with asymptotic results, however.

Example 1.1. An example of a global exponential mapping is the following. Let $\mathcal{M} = \mathcal{S}^1 = \{(x_1, x_2) \in \boldsymbol{R}^2 : x_1^2+x_2^2 = 1\}$. The exponential mapping at the point (1, 0) is

$$h \to \exp(h) = (\cos h, \sin h),\ h \in \boldsymbol{R}, \qquad \dots \quad (1.1)$$

which might of course be identified with $\exp(ih)$, i imaginary unit. Here we identify the tangent space $\boldsymbol{R}\times\{0\}$ in the usual way with $\boldsymbol{R}$. Given a random variable $U \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$, $\exp(U)$ defines a probability measure which is called the wrapped normal distribution with parameters $\mu \in \boldsymbol{R}$ and $\sigma^2 \in (0, \infty)$. In Mardia (1972) the definition is essentially the same, but the relation with the exponential mapping is not mentioned. It is clear that in polar coordinates $(\cos\varphi, \sin\varphi)$, $\varphi \in (0, 2\pi]$, the density of this wrapped normal distribution is given by

$$\tilde{f}_{\mu,\sigma}(\varphi) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{k=-\infty}^{\infty} \exp\left[-\frac{1}{2}\left[\frac{\varphi-\mu+2\pi k}{\sigma}\right]^2\right]$$

$$= (2\pi)^{-1}\left[1+2\sum_{k=1}^{\infty}\exp\left(-\frac{1}{2}(k\sigma)^2\right)\cos k(\varphi-\mu)\right],\ 0 < \varphi \leqslant 2\pi. \quad \ldots \quad (1.2)$$

For this useful equality see Mardia (1972, formula (3.4.30)).

Now suppose that $\Phi_1, \ldots, \Phi_n$ are i.i.d. real valued random variables with finite expectation μ and finite non-zero variance σ^2. Hence Exp $(\Phi_i) = (\cos\Phi_i, \sin\Phi_i)$ are i.i.d. random variables in $\mathcal{S}^1$. To determine a suitable sum with a limiting law on $\mathcal{S}^1$ it seems natural to first take the usual partial sum of the Φ_i in the tangent space, and then map back onto $\mathcal{S}^1$ using Exp. Hence we consider

$$\exp\left[\frac{\sum_{i=1}^{n}(\Phi_i-\mu)}{n^{\frac{1}{2}}\sigma}\right] = \left(\cos\left[\frac{\sum_{i=1}^{n}(\Phi_i-\mu)}{n^{\frac{1}{2}}\sigma}\right],\ \sin\left[\frac{\sum_{i=1}^{n}(\Phi_i-\mu)}{n^{\frac{1}{2}}\sigma}\right]\right), \quad \ldots \quad (1.3)$$

which is equivalent to the random variable considered in Mardia (1972, Section 4.3.2b). It is clear that the random variables in (1.3) converge in distribution to the wrapped normal distribution with $\mu = 0$ and $\sigma = 1$.

More generally the special orthogonal group admits an onto exponential mapping. In principle for such manifolds probability distributions and weak convergence may be defined with the aid of the usual theory for random vectors in linear spaces.

1.2. *Mean location.* Let $\mathcal{M} \subset \boldsymbol{R}^k$ be a smooth manifold of dimension $1 \leqslant m \leqslant k$, let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and $S : \mathcal{X} \to \mathcal{M}$ a random variable in the manifold. The *mean location* of S is defined to be the point $M(S) \in \mathcal{M}$, unique by assumption, satisfying

$$E\|S-M(S)\|^2 = \inf_{p\in\mathcal{M}} E\|S-p\|^2. \quad \ldots \quad (1.4)$$

Hence $M(S)$ is the projection of $E(S)$ on $\mathcal{M}$. Uniqueness is typically satisfied when the law of S on $\mathcal{M}$ is sufficiently concentrated.

The expectation of $S = (S_1, \ldots, S_k)$ is defined in the usual way be $E(\mathrm{S}) = (E(S_1), \ldots, E(S_k))$ and hence satisfies

$$E\|S-E(S)\|^2 = \inf_{\boldsymbol{x}\in\boldsymbol{R}^k} E\|\boldsymbol{S}-\boldsymbol{x}\|^2. \quad \ldots \quad (1.5)$$

It is clear that $M(S) = E(S)$ when $\mathcal{M}$ is a convex set.

Example 1.2. Let again $\mathscr{M} = \mathscr{S}^1 \subset \boldsymbol{R}^2$ and $S : \mathscr{X} \to \mathscr{S}^1$ be given by $S = (\cos \Phi, \sin \Phi)$ for $\Phi : \mathscr{X} \to \boldsymbol{R}$. Then Mardia (1972, formula (3.3.7)) defines the mean direction as $\mu \in (0, 2\pi]$ and $r \geqslant 0$ are given by $(E \cos \Phi, E \sin \Phi) = r(\cos \mu, \sin \mu)$. This yields

$$((E \cos \Phi)^2 + (E \sin \Phi)^2)^{-\frac{1}{2}} (E \cos \Phi, E \sin \Phi) \in \mathscr{S}^1, \qquad \ldots \quad (1.6)$$

as the corresponding point on the circle (assuming $r > 0$). It is easily seen that (1.6) minimizes (1.4), so that the point coincides with the mean location.

Let us introduce the modified Bessel function

$$I_m(\kappa) = \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\kappa)^{2k+m}}{\Gamma(m+k+1)\,\Gamma(k+1)}, \quad \kappa \in \boldsymbol{R}, m \in \mathscr{Z}, \qquad \ldots \quad (1.7)$$

where Γ denotes the gamma-function and $\mathscr{Z}$ the set of all integers. It is shown in Mardia (1972) that unimodal symmetric distributions on the circle have the symmetry point at the mode as mean direction. Hence it follows that for both the wrapped normal distribution in (1.2) and the von Mises distribution with density

$$f_{\mu,\kappa}(\varphi) = \frac{\exp(\kappa \cos(\varphi - \mu))}{2\pi\, I_0(\kappa)}, \; 0 < \varphi \leqslant 2\pi, \qquad \ldots \quad (1.8)$$

when $\kappa > 0$ and $0 < \mu \leqslant 2\pi$, the mean direction is μ and hence $(\cos \mu, \sin \mu)$ is the mean location.

1.3. *Divergence*. Let $T_{M(S)}\mathscr{M}$ be the tangent space to $\mathscr{M}$ at $M(S)$, as usual identified with an m-dimensional subspace of $\boldsymbol{R}^k$. For arbitrary $h \in T_{M(S)}\mathscr{M}$ we define the *divergence* of S in the direction h by

$$D_h(S) = \mathrm{Var}(h, S) = \mathrm{Var} \sum_{j=1}^{k} h_j S_j, \qquad \ldots \quad (1.9)$$

where $(.\,,.)$ denotes inner product in $\boldsymbol{R}^d$ for any dimension $d \in \boldsymbol{N}$.

Since $E(S) - M(S)$ is by definition of $M(S)$ perpendicular to the tangent space $T_{M(S)}\mathscr{M}$ we may write alternatively

$$D_h(S) = E(S - M(S), h)^2. \qquad \ldots \quad (1.10)$$

Values of S for which $S - M(S)$ is almost perpendicular to h don't contribute very much to the value of the divergence. Such points may be found on compact manifolds without boundary, like $\mathscr{S}^{k-1}$, in a neighborhood of the point of $\mathscr{M}$ which maximizes distance to $M(S)$. But even then it is a resonable measure for dispersion when the law of S is concentrated about $M(S)$ as such values of S are not arbitrarily lose to $M(S)$.

Example 1.3. In the notation of Example 1.2 it follows by elementary geometry that $(h = 1)$

$$D_1(S) = E \sin^2 (\Phi-\mu), \qquad \ldots \quad (1.11)$$

where $\mu \epsilon (0, 2\pi]$ is determined by $M(S) = (\cos \mu, \sin \mu)$. In Mardia (1972, formula (3.3.9)) the circular variance is defined by

$$V_0(S) = 1-E \cos (\Phi-\mu), \qquad \ldots \quad (1.12)$$

and, apart from conditioning, the divergence (Mardia, 1972, formula (5.1.6)) as

$$\text{div}(S, \mu) = \frac{E \sin^2 (\Phi-\mu)\ 2}{(E \cos (\Phi-\mu))} \qquad \ldots \quad (1.13)$$

Let us now assume that S has the distribution with density in (1.8). It follows from Mardia (1972, formulas (5.1.9) and (3.4.47)) that

$$D_1(S) = I_1(\kappa)/(\kappa I_0(\kappa)), \qquad \ldots \quad (1.14)$$

and $V_0(S) = 1-I_1(\kappa)/I_0(\kappa)$ respectively.

When S has the wrapped normal distribution with density (1.2) we have, writing $\exp\left(-\frac{1}{2}\ \sigma^2\right) = \rho$,

$$D_1(S) = 1-\int_0^{2\pi} \cos^2 (\varphi-\mu)\tilde{f}_{\mu,\sigma}(\varphi)d(\varphi)$$

$$= 1-\int_0^{2\pi} \frac{\cos^2 \varphi}{2\pi} \left[1+2 \sum_{k=1}^{\infty} \rho^{k^2} \cos k\varphi\right] d\varphi$$

$$= 1-\frac{1}{2} -(1/\pi)\rho^4(1/2)\ \pi = \frac{1}{2}\ (1-\rho^4). \qquad \ldots \quad (1.15)$$

According to Mardia (1972, formula (3.4.32)) the circular variance equals $V_0(S) = 1-\rho$.

2. The inequality

In this section $(\mathscr{X}, \mathscr{A}, \{P_\theta: \theta \epsilon \Theta\})$ is a statistical model with $\{P_\theta : \theta\epsilon\Theta\} \ll \mu$, for some σ-finite measure μ on $\mathscr{A}$. We set

$$f_\theta = \frac{dP_\theta}{d\mu}\ ; l_\theta = \log(f_\theta\ ),\ all\ \theta \epsilon \Theta. \qquad \ldots \quad (2.1)$$

(assuming that $f_\theta > 0$ for all $\theta \epsilon \Theta$). Furthermore, the parameter space is a smooth (i.e. of class C^∞) submanifold of dimension n in $\boldsymbol{R}^p$: $1 \leqslant n \leqslant p$. Throughout a $t \in \Theta$ is kept fixed. As usual $C^\infty(t)$ denotes the set of C^∞-functions which have some open neighbourhood of t in Θ as domain. For each $x \in \mathscr{X}$ the function $\theta \to l_\theta(x)$ is supposed to be in $C^\infty(t)$. When $g \in C^\infty(t)$ we will denote its derivative in the direction $\xi \in T_s\Theta$ (for s in the domain of g) by $\nabla_\xi g_\theta$. Finally, for a C^∞ map $F : U \to \boldsymbol{R}^k$($U$ is an open neighbourhood of t on Θ) $\nabla_\xi F_\theta$ denotes the vector of derivatives in the direction $\xi \in T_s\Theta$, $s \in U$.

Let $S : \mathscr{X} \to \mathscr{M}$ be a statistic where $\mathscr{M}$ is an m-dimensional smooth submanifold of $\boldsymbol{R}^k$. Assume that the maps $\theta \to E_\theta(S)$, $\theta \to M_\theta(S)$ and $\theta \to \Sigma_\theta(S)$ (the covariance matrix mapping) exist and are componentwise in $C^\infty(t)$ Notice the local character of the regularity conditions introduced in this section. Assume furthermore

$$I_{\xi,\xi}(t) = E_t(\nabla_\xi l_\theta)^2 > 0, \ all \ \xi \in T_t\Theta, \qquad \dots \quad (2.2)$$

and for some open neighbourhood U_t of t let $h : U_t \to \boldsymbol{R}^k : \theta \to h_\theta$ be a C^∞-vectorfield with $h_\theta \in T_{M_\theta(S)}\mathscr{M}$ for all $\theta \in \Theta$ Notice that h may be chosen so as to assume any vector from $T_{M_t(S)}\mathscr{M}$ at $\theta = t$. Such a map h exists even globally.

Under the conditions mentioned above, we have

$$D_{t,h_t}(S) \geqslant \frac{(\nabla_\xi E_\theta(S), h_t)^2}{I_{\xi,\xi}(lt)}, \qquad \dots \quad (2.3)$$

for all $\xi \in T_t\Theta$ ($= T_tU_t$, since $U_t \subset \Theta$ open). The classical proof where Θ is an open subset of Euclidean space needs only a minor modification to suit the present more general assumption that Θ is a smooth submanifold. We have seen in (1.10) that we may replace the l.h.s. of (2.3) by $E_t(S - M_t(S), h_t)^2$ which is indeed the natural quantity to look at since it measures the dispersion about the mean location on the manifold. We may also rewrite the r.h.s. in (2.3) in such a way that some of the geometry of the manifold $\mathscr{M}$ is reflected (see the comments after Theorem 2.1). Using again the fact $E_\theta(S) - M_\theta(S)$ is normal to $T_{M_\theta(S)}\mathscr{M}$, for all $\theta \in \Theta$, we obtain

$$\begin{aligned}(\nabla_\xi E_\theta(S), h_t) &= \nabla_\xi(E_\theta(S), h_\theta) - (E_t(S), \nabla_\xi h_\theta) \\ &= \nabla_\xi(M_\theta(S), h_\theta) - (E_t(S), \nabla_\xi h_\theta) \\ &= (\nabla_\xi M_\theta(S), h_t) + (M_t(S) - E_t(S), \nabla_\xi h_\theta). \qquad \dots \quad (2.4)\end{aligned}$$

Summarizing we have proved the following result.

Theorem 2.1. *Under the conditions mentioned above, we have*

$$E_t(S-M_t(S),\, h_t)^2 \geqslant \frac{\{(\nabla_\xi M_\theta(S),\, h_t)+(M_t(S)-E_t(S),\, \nabla_\xi\, h_\theta)\}^2}{I_{\xi,\xi}(t)} \quad \ldots \quad (2.5)$$

for all $\xi \in T_t\, \Theta$.

Now the differential geometric relevance of the term $(M_t(S)-E_t(S),\, \nabla_\xi h_\theta)$ (which still contains the expectation of S) should be noted. Suppose namely that dim $(\mathcal{M}) = k-1$ ($\mathcal{M}$ has unit codimension), and let $\xi,\, \zeta \in T_{M_t(S)}\, \mathcal{M}$. Furthermore, let X be a tangent vectorfield on $\mathcal{M}$ such that $X(M_t(S)) = \zeta$. As $N_t = \|M_t(S)-E_t(S)\|^{-1}\, (M_t(S)-E_t(S))$ is, by definition of $M_t(S)$, a unit vector normal to $T_{M_t(S)}\, \mathcal{M}$ we have

$$(M_t(S)-E_t(S),\, \nabla_\xi X) = -\|M_t(S)-E_t(S)\|(L(\xi),\, \zeta), \quad \ldots \quad (2.6)$$

where the linear map $L = L_t : T_{M_t(S)}\, \mathcal{M} \rightarrow T_{M_t(S)}\, \mathcal{M}$ is known as the Weingarten map in differential geometry (see e.g. Spivak (1999, p. 101)).

For a hypersurface $\mathcal{M}$ of Euclidean space $\boldsymbol{R}^k$ the Weingarten map (or shape operator ; see e.g. Boothby (1986, p. 368) or Spivak (1999, p. 98) is defined by

$$L_N(X) = \nabla_X N, \textit{ for every tangent vector of } \mathcal{M}, \quad \ldots \quad (2.7)$$

where N is a C^∞-vectorfield of unit length normal to $\mathcal{M}$. In Spivak (1975, p. 71 and subsequent discussion) it is proved that this map together with the Riemannian structure determines the hypersurface $\mathcal{M}$ up to a proper rigid motion, when $\mathcal{M}$ is oriented. For arbitrary codimension the general theory is less satisfactory. It turns out, however, that in general when $I+(1-\|M_t(S)-E\,(S)\|)L$ (the map L defined by formula (2.6)) is singular, $E_t(S)$ coincides with a centre of curvature of $\mathcal{M}$ and thus $\mathcal{M}_t\, (S)$ is ill-determined.

Returning to the topic of this subsection, in order to maximize the right hand side of (2.3) let us introduce the following maps on $T_t\, \Theta$

$$g_t(\xi) = (\nabla_\xi M_\theta(S),\, h_t)+(M_t(S)-E_t(S),\, \nabla_\xi h_\theta), \quad \ldots \quad (2.8)$$

$$I_{\xi,\eta}(t) = E_t(\nabla_\xi l_\theta)\,(\nabla_\eta l_\theta)\ all\ \xi,\, \eta \in T_t\Theta. \quad \ldots \quad (2.9)$$

Notice that $g_t \in T_t^*\Theta$. Furthermore $(\xi,\, \eta) \rightarrow I_{\xi,\eta}(t)$ induces a natural linear map $\tilde{I}_t : T_t\Theta \rightarrow T_t^*\Theta$ via

$$\tilde{I}_t(\xi)\,(\eta) = I_{\xi,\eta}\,(t) : all\ \xi,\, \eta \in T_t\Theta. \quad \ldots \quad (2.10)$$

As $I_{\xi,\xi}(t) > 0$ for all $\xi \in T_t\Theta$ the map $\tilde{I}_t$ is invertible.

Theorem 2.2. *Under the assumptions of this section, we have*

$$D_{t,h_t}(S) = E_t(S-M_t(S),\, h_t)^2 \geqslant g_t(\tilde{I}_t^{-1}(g_t)). \qquad \dots \quad (2.11)$$

Proof. Let $T_t\Theta$ be topologised by the norm topology (all norms on $T_t\Theta$ being equivalent). With the notation just introduced, Theorem 2.1 implies

$$D_{t,h_t}(S) = E_t(S-M_t(S),\, h_t)^2 \geqslant \sup_{\xi \in T_t\Theta \setminus \{0\}} \frac{(g_t(\xi))^2}{I_{\xi,\xi}(t)}. \qquad \dots \quad (2.12)$$

As both maps appearing on the right in (2.12) are quadratic in ξ, the differentiability of the map $F : T_t\Theta\setminus\{0\} \to \boldsymbol{R} : \xi \to [g_t(\xi)]^2/I_{\xi,\xi}(t)$ is guaranteed, as well as the assumption of a global maximum at $T_t\Theta\setminus\{0\}$ (restrict F to a unit sphere). Differentiating F under the constraint $I_t =$ constant (i.e. applying the Lagrange-method) shows that F assumes $g_t(\tilde{I}_t^{-1}(g_t))$ as a global maximum at $\xi = \tilde{I}^{-1}(g_t)$. Q.E.D.

As a special case of Theorem 2.2 take $\Theta \subset \boldsymbol{R}^p$ to be open and non void. Let J_t be the classical Fisher information matrix. Then J_t is the matrix of $(\xi, \eta) \to I_{\xi,\eta}(t)$ on the natural basis of $\boldsymbol{R}^p$. Let G_t be the matrix, relative to the same basis, of $(\xi, \eta) \to g_t(\xi)g_t(\eta)$. In this interpretation we have

$$g_t(\tilde{I}_t^{-1}(g_t)) = \textit{largest eigenvalue of } J_t^{-1/2}\, G_t J_t^{-1/2}. \qquad \dots \quad (2.13)$$

Hence

$$D_{t,h_t}(S) = E_t(S-M_t(S),\, h_t)^2 \geqslant \textit{largest eigenvalue of } J_t^{-1/2}\, G_t J_t^{-1/2}. \qquad \dots \quad (2.14)$$

If, in addition, we simply take $\mathscr{M} = \boldsymbol{R}^k$ then for all θ in Θ, $M_\theta(S) = E_\theta(S)$ as observed before and the tangent space at any $E_\theta(S)$ may now be identified with $\boldsymbol{R}^k$ itself. Hence for the smooth vectorfield we may now take the constant field $h_\theta = h$, $\theta \in \Theta$, for some fixed $h \in \boldsymbol{R}^k$.

As has been noted in the introduction, the sphere and the special orthogonal group are important special cases. Since the treatment of the latter is rather technical it will be postponed to the next section. This section will be concluded with the somewhat simpler case of the sphere.

Example 2.1. Let $\Theta = \mathscr{M} = \mathscr{S}^{k-1}$, the unit sphere is $\boldsymbol{R}^k$. Let the parameter $a \in \mathscr{S}^{k-1}$ be fixed. For (1.4) we find

$$E_a\|S-M_a(S)\|^2 = 2-2 \max_{x \in S^{k-1}} (E_a(S), x). \qquad \dots \quad (2.15)$$

Throughout we assume that $E_x(S) \neq 0$ for all x in $\mathscr{S}^{k-1}$. Since $M_a(S)$ is the projection on the sphere of $E_a(S)$ it is immediate that

$$M_a(S) = \|E_a(S)\|^{-1}\, E_a(S). \qquad \dots \quad (2.16)$$

Now let X and Z be vectors in $T_{M_a(S)}\mathcal{S}^{k-1} = [E_a(S)]^{\perp}$ and let $Y: \mathcal{S}^{k-1} \to \boldsymbol{R}^k$ be a C^∞-vectorfield with $Y_a = X$ and in general $Y_b \in T_{M_b(S)}\mathcal{S}^{k-1}$, for all b in $\mathcal{S}^{k-1}$. In order to specialise the information inequality (2.11) to the case of the sphere first note that (see definition (1.9)) :

$$D_{a,X} = E_a(S-M_a(S), X)^2 = \mathrm{Var}_a\,(S, X). \qquad \dots \quad (2.17)$$

Next consider the linear map g_a as in (2.13) (take $t = a$) :

$$g_a(Z) = (\nabla_Z M_x(S), X) + (M_a(S) - E_a(S), \nabla_Z Y). \qquad \dots \quad (2.18)$$

Let φ denote the C^∞ map $x \to M_x(S)$ on $\mathcal{S}^{k-1}$ and φ_* the associated Jacobian map. Then $\nabla_Z M_x(S) = \varphi_*(Z)$. Let φ^* be the dual map of φ_*. After identifying $T_b^*\mathcal{S}^{k-1} = T_b\mathcal{S}^{k-1}$ through the usual inner product on $\boldsymbol{R}^k$ we get $(\nabla_Z M_x(S), X) = (Z, \varphi^*(X))$. Now let L be the Weingarten map as introduced in (2.6) (in fact L is the identity map). Since $\|M_a(S)\| = 1$ and $\|E_a(S)\| \leqslant 1$ the vector $M_a(S) - E_a(S)$ is an outward pointing normal to $\mathcal{S}^{k-1}$ in $M_a(S)$ so that equality (2.6) implies

$$\begin{aligned}(M_a(S) - E_a(S), \nabla_Z Y) &= -\|M_a(S) - E_a(S)\|(L(Z), X) \\ &= (\|E_a(S)\| - 1)\,(Z, X). \qquad \dots \quad (2.19)\end{aligned}$$

Lastly, using the identification mentioned above, the substitution of the above results in (2.18) and the combination of (2.17) with the general inequality (2.11) gives the information inequality

$$\begin{aligned}&\mathrm{Var}_a(S, X) \geqslant g_a^t \tilde{I}_a^{-1} g_a, \text{ with} \\ &\qquad g_a = \varphi^*(X) + (\|E_a(S)\| - 1)X, \text{ all } X \text{ in } T_{M_a(S)}\mathcal{S}^{k-1}, \quad \dots \quad (2.20)\end{aligned}$$

in the case of the sphere.

This formula becomes much simpler when $M_x(S) = x$ for all $x \in \mathcal{S}^{k-1}$. Statistically this is a kind of unbiasedness of S since its mean location under x coincides with the parameter x. Now φ^* is the identity so that we simply obtain

$$g_a = \|E_a(S)\|.\,X, \text{ all } X \text{ in } T_{M_a(S)}\,\mathcal{S}^{k-1}. \qquad \dots \quad (2.21)$$

3. Examples : $\mathcal{SO}(k)$ and some of its subgroups

In this section we will explicitly compute the mean location and the lower bound for random variables in the orthogonal group. We will also briefly consider the problem of attainment of the lower bound. Let $\mathcal{M}_k$ be the set of $k \times k$-matrices $M = (M_{qr})_{q,r \in \{1, \dots, k\}}$ over the real numbers. We will identify (e.g. topologically) $\mathcal{M}_k$ with $\boldsymbol{R}^{k\times k}$. The transpose of $M = (M_{qr})$

is defined as $M^t = (M^t_{qr}) = (M_{rq})$, and the trace of M by $\mathrm{tr}M = \sum_{q=1}^{k} M_{qq}$. Given any two matrices $M_1, M_2 \in \mathcal{M}_k$ we have

$$\mathrm{tr}(M_1+M_2) = \mathrm{tr}M_1+\mathrm{tr}M_2,\ \mathrm{tr}M_1M_2 = \mathrm{tr}M_2M_1, \qquad \dots \quad (3.1)$$

and the important relation with the inner product

$$(M_1, M_2) = \mathrm{tr}M_1M_2. \qquad \dots \quad (3.2)$$

For any smooth manifold $\mathcal{M} \subset \mathcal{M}_k$ the metric will always be derived from (3.2). Let I denote the identity matrix.

The smooth manifold that we will be concerned with has, moreover, a group structure and is called the *special orthogonal group*

$$\mathcal{SO}(k) = \{O \in \mathcal{M}_k : O^t O = I, \textit{ and } \det O = 1\}, \qquad \dots \quad (3.3)$$

consisting of all orthonormal matrices with determinant $+1$. Let us also introduce the *linear subspace* of all *skew-symmetric* matrices

$$\boldsymbol{H}_k = \{H \in \mathcal{M}_k : H = -H^t\}. \qquad \dots \quad (3.4)$$

The dimension of this subspace is $\frac{1}{2}(k-1)k$ and we have the useful relation

$$\mathrm{tr}H = 0,\ H \in \boldsymbol{H}_k. \qquad \dots \quad (3.5)$$

We will make use of the well-known fact (see e.g. Curtis, 1985) that

$$T_O\,\mathcal{SO}(k) = \{OH : H \in \boldsymbol{H}_k\} = \{HO : H \in \boldsymbol{H}_k\},\ O \in \mathcal{SO}(k). \qquad \dots \quad (3.6)$$

It will be tacitly understood that the conditions mentioned in the beginning of Section 2.1 are satisfied and that $\tilde{I}_t$ is invertible. Hence the expression on the right in (2.3) may be replaced by its maximum over all $\xi \in \boldsymbol{R}^p$.

3.1. *Computation mean location.* Let $S : \mathcal{X} \to \mathcal{SO}(k)$ be a statistic such that all the regularity conditions of Section 2 hold. Throughout the parameter $\theta \in \Theta$ is kept fixed. Set $A_\theta = E_\theta(S)$ and suppose that $\det(A_\theta) \neq 0$. Furthermore define $\psi : \mathcal{SO}(k) \to \boldsymbol{R}$ by $\psi(x) = (A_\theta, x) : x \in \mathcal{SO}(k)$. Then we have

$$E_\theta\|S-x\|^2 = 2(k-\psi(x)),\ \textit{all } x \in \mathcal{SO}(k). \qquad \dots \quad (3.7)$$

Hence, as $\mathcal{SO}(k)$ is compact we have

$$\inf_{x \in \mathcal{SO}(k)} E_\theta\|S-x\|^2 = 2(k-\max_{x\in\mathcal{SO}(k)} \psi(x)). \qquad \dots \quad (3.8)$$

Since $\psi \in C^\infty(\mathcal{SO}(k))$, determining $M_\theta(S)$ apparently amounts to calculating the stationary points of ψ (which exist). Now let $\boldsymbol{M}_k$ denote the set of symmetric $k\times k$-matrices in $\boldsymbol{R}^{k\times k}$. Then $\boldsymbol{M}_k = \boldsymbol{H}_k^\perp$.

Suppose $x_0 \in \mathcal{SO}(k)$ is a critical point of ψ. As for each $H \in \boldsymbol{H}_k$, $\gamma_{x_0,H} : \boldsymbol{R} \to \mathcal{SO}(k)$ such that $s \to x_0 \exp(sH)$ is a smooth curve through x_0 we have : $\gamma'_{x_0,H}(0) = x_0 H$ hence

$$(x_0^t A_\theta, H) = (A_\theta, x_0 H) = (\psi \circ \gamma_{x_0,H})'(0) = 0, \text{ all } H \in \boldsymbol{H}_k. \quad \ldots \quad (3.9)$$

Apparently $x_0 = (A_\theta^t)^{-1} M$, for some M in $\boldsymbol{M}_k$. Since $x_0 \in \mathcal{SO}(k)$, (3.9) implies that for $(A_\theta^t)^{-1}M$ to be in $\mathcal{SO}(k)$ we must have :

$$\det(M) = \det(A_\theta) \text{ and } M^2 = A_\theta^t A_\theta. \quad \ldots \quad (3.10)$$

Conversely when $M \in \boldsymbol{M}_k$ is such that (3.10) holds, one easily shows that $(A_\theta^t)^{-1} M$ is a critical point of ψ.

Let $\sum_{i=1}^{r} \lambda_i P_i$ with $\lambda_1 > \ldots > \lambda_r > 0$ be the spectral decomposition of the positive definite symmetric matrix $A_\theta^t A_\theta$. As a direct consequence of the uniqueness of such a decomposition we get : $M \in \boldsymbol{M}_k$ satisfies (3.10) iff

$$M = M_a := \sum_{i=1}^{r} a_i \sqrt{\lambda_i} P_i, \quad \ldots \quad (3.11)$$

for some $a = (a_1, \ldots, a_r) \in \{-1, 1\}^k$ such that $\prod_{i=1}^{r} a = \mathrm{sgn}(\det A_\theta)$.

Now suppose $M = M_a$ is as in (3.11). Then

$$\psi((A_\theta^t)^{-1} M) = (A_\theta, (A_\theta^t)^{-1} M) = \mathrm{tr} M_a = \sum_{i=1}^{r} a_i \sqrt{\lambda_i}. \quad \ldots \quad (3.12)$$

Therefore when a is such that $a \equiv 1$ for all i (except possibly for $i = r$) then ψ assumes its global maximum at $(A_\theta^t)^{-1} M_a$. Hence for the mean location we finally find

$$M_\theta(S) \begin{cases} = (E_\theta^t(S))^{-1} \sum_{i=1}^{r} \sqrt{\lambda_i} P_i = (E_\theta^t(S))^{-1} |E_\theta(S)| \text{ if } \det(E_\theta(S)) > 0, \\ = (E_\theta^t(S))^{-1} \left(\sum_{l=1}^{r-1} \sqrt{\lambda_i} P_i - \sqrt{\lambda_r}\, P_r \right) \\ = (E_\theta^t(S))^{-1} (|E_\theta^t(S)| - 2\sqrt{\lambda_r} P_r) \\ = (E_\theta^t(S))^{-1} |E_\theta^t(S)| - 2\lambda_r^{-1/2} E_\theta(S).\ P_r \text{ if } \det(E_\theta(S)) < 0, \end{cases} \quad \ldots \quad (3.13)$$

where $|E_\theta(S)| = \sqrt{\{E_\theta^t(S).E(S)\}^{i}}$ is the unique positive definite square root of $E_\theta^t(S).\, E_\theta(S)$. Notice that $M_\theta(S) = E_\theta(S)$ iff $E_\theta(S) \in \mathcal{SO}(k)$.

Example 3.1. As a special case consider the group $\mathcal{SO}(2)$. Let $S : \mathcal{X} \to \mathcal{SO}(2)$, say $S = \begin{pmatrix} S_1 & -S_2 \\ S_2 & S_1 \end{pmatrix}$ with $S_1^2 + S_2^2 \equiv 1$. Then $\det E_\theta(S) = E_\theta(S_1)^2 + E_\theta(S_2)^2 > 0$. Using (3.13) we get

$$
\begin{aligned}
M_\theta(S) &= E_\theta^{t}(S)^{-1} \,|\, E_\theta(S) \,| \\
&= [E_\theta^2(S_1)+E_\theta^2(S_2)]^{-1} \begin{pmatrix} E_\theta(S_1) & -E_\theta(S_2) \\ E_\theta(S_2) & E_\theta(S_1) \end{pmatrix} |\, E_\theta(S) \,| \\
&= [E_\theta^2(S_1)+E_\theta^2(S_2)]^{-1} \begin{pmatrix} E_\theta(S_1) & -E_\theta(S_2) \\ E_\theta(S_2) & E_\theta(S_1) \end{pmatrix} \\
&\qquad \left[\begin{pmatrix} E_\theta(S_1) & E_\theta(S_2) \\ -E_\theta(S_2) & E_\theta(S_1) \end{pmatrix} \begin{pmatrix} E_\theta(S_1) & -E_\theta(S_2) \\ E_\theta(S_2) & E_\theta(S_1) \end{pmatrix} \right]^{1/2} \\
&= [E_\theta^2(S_1)+E_\theta^2(S_2)]^{-1/2} \begin{pmatrix} E_\theta(S_1) & -E_\theta(S_2) \\ E_\theta(S_2) & E_\theta(S_1) \end{pmatrix} \sqrt{Id} \\
&= [E_\theta^2(S_1)+E_\theta^2(S_2)]^{-1/2}\, E_\theta(S) = \sqrt{2}\|E_\theta(S)\|^{-1}\, E_\theta(S), \qquad \ldots \quad (3.14)
\end{aligned}
$$

cf. Example 1.2 ($\mathcal{SO}(2)$ is a circle of radius $\sqrt{2}$ in $\boldsymbol{R}^4$).

3.2. *Computation lower bound.* Let $H \in \boldsymbol{H}_k \setminus \{0\}$. As vectorfield $\theta \to h_\theta$ we may now take $\theta \to M_\theta(S).H$ (which is nonvanishing globally). For $\xi \in T_t\Theta$ (notation of Section 2) we have

$$
\begin{aligned}
g_t(\xi) &= (\nabla_\xi M_\theta(S),\, h_t)+(M_t(S)-E_t(S),\, \nabla_\xi M_\theta(S).H) \\
&= (\nabla_\xi M_\theta(S),\, M_t(S)\, H)-(\nabla_\xi M_\theta(S),\, M_t(S).H-E_t(S)\,.\,H) \\
&= (\nabla_\xi M_\theta(S),\, E_t(S)\,.\,H). \qquad \ldots \quad (3.15)
\end{aligned}
$$

Finally notice that equality (2.7) shows that in order to calculate the lower bound in this context we need not consider two cases, as in (3.13).

Example 3.2. The unit circle in $\boldsymbol{R}^2$ may be identified with $\mathcal{SO}$ (2), hence we write $S = \begin{pmatrix} S_1 & -S_2 \\ S_2 & S_1 \end{pmatrix}$ with $S_1^2+S_2^2 \equiv 1$. Now let $t \in \Theta$ be fixed with $\Theta \subset \boldsymbol{R}$ open, non void. Let $H = \begin{pmatrix} 0 & h \\ -h & 0 \end{pmatrix}$ with $h \in \boldsymbol{R}\setminus\{0\}$ fixed. Using subsection 3.2, we find

$$
\begin{aligned}
g_t &= \sqrt{2}((d/d\theta)_t\, \|E_\theta\,(S)\|^{-1}\, E_\theta\,(S),\, E_t\,(S)H) \\
&= -2h\sqrt{2}\|\, E_t\,(S)\|^{-1}\, [E_t\,(S_2)(d/d\theta)_t\, E_\theta(S_1)-E_t(S_1)\,(d/d\theta)_t\, E_\theta(S_2)], \ldots \quad (3.17)
\end{aligned}
$$

where $(d/d\theta)_t$ denotes ordinary differentiation w.r.t. the real variable θ at $\theta = t$. For the sake of brevity set $u_t = E_t(S_2)\,(d/d\theta)_t\, E_\theta(S_1)-E_t(S_1)(d/d\theta)_t$ $E_\theta(S_2)$ and let J_t be the classical Fisher-information. Then after some obvious identifications we find for the lower bound (see (2.12))

$$
g_t(\widetilde{I_t}^{-1}\,(g_t)) = 8h^2\, \|E_t\,(S)\|^{-2} u_t^2/J_t. \qquad \ldots \quad (3.18)
$$

Furthermore the divergence term in (2.12) reduces to

$$D_{t,h_t}(S) = \mathrm{Var}_t\,(S,\, M_t\,(S)H)$$
$$= 2h^2\, \|E_t\,(S)\|^{-2}\, \mathrm{Var}_t\,(S_1E_t\,(S_2) - S_2E_t\,(S_1)). \qquad \ldots \quad (3.19)$$

Combining (3.18) and (3.19) gives for this case the desired Cramér-Rao type inequality

$$\mathrm{Var}_t\,(S_1E_t\,(S_2) - S_2E_t\,(S_1)) \geqslant 4u_t^2/J_t, \qquad \ldots \quad (3.20)$$

with u_t as defined above.

In order to compare (3.20) with the corresponding results in Mardia (1972) introduce a random variable $\Phi : \mathscr{X} \to \boldsymbol{R}$ such that $S_1 = \cos\Phi$; $S_2 = \sin\Phi$. Then there exists a smooth function $\psi : \Theta \to (0, 2\pi]$ such that for all $\theta \in \Theta$: $E_\theta \cos\Phi = r(\theta)\cos\psi(\theta)$ and $E_\theta \sin\Phi = r(\theta)\sin\psi(\theta)$ where $r(\theta) = (E_\theta^2 \cos\Phi + E_\theta^2 \sin\Phi)^{-1/2}$. Making these substitutions in (3.20) gives

$$E_t \sin^2(\Phi - \psi(t)) \geqslant [\psi'(t)E_t \cos(\Phi - \psi(t))]^2 .\ J_t^{-1}, \qquad \ldots \quad (3.21)$$

which apart from the conditioning is essentially the same inequality as Mardia (1972, formula 5.1.5).

3.3. *Attainment Lower Bound.* Using equation (2.6) it is possible to give a non-trivial example in which attainment occurs. Let $\mathscr{X} = \mathscr{M} = \mathcal{SO}(k)$ Let μ denote Haar-probability measure on $\mathcal{SO}(k)$. As family of densities on $\mathscr{X}$ consider $\{f_A : A \in \mathcal{SO}(k)\}$, with

$$f_A(x) = C_\sigma \exp\left(\frac{2}{\sigma}(x, A)\right),\ x \in \mathcal{SO}(k),\ \sigma > 0 \text{ fixed}, \qquad \ldots \quad (3.22)$$

and C_σ is a norming constant (independent of A). We will show that the statistic $S(x) = x$ (i.e. the identity map on $\mathcal{SO}(k)$) attains the lower bound.

First we prove that for all $A \in \mathcal{SO}(k)$: $M_A(S) = A$. To accomplish this, set $M = \int_{\mathcal{SO}(k)} x \exp\left[\frac{2}{\sigma}\,\mathrm{tr}(x)\right] d\mu(x)$(Pettis-integration). As conjugation is measure preservaing M commutes with every matrix in $\mathcal{SO}(k)$, hence $M = \alpha I$ for some constant α. We claim that $\alpha > 0$ (i.e. that M is positive definite and symmetric). Remark that

$$k\,.\,\alpha = \mathrm{tr}(M) = \int_{\mathcal{SO}(k)} \mathrm{tr}(x)\exp\left[\frac{2}{\sigma}\,\mathrm{tr}(x)\right] d\mu(x). \qquad \ldots \quad (3.23)$$

Here, the integrand has the following uniformly converging Taylor series :

$$\mathrm{tr}(x)\,\exp\left[\frac{2}{\sigma}\,\mathrm{tr}(x)\right] = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{2}{\sigma}\right)^n (\mathrm{tr}(x))^{n+1}. \qquad \ldots \quad (3.24)$$

Now $\int_{\mathcal{SO}(k)} (\mathrm{tr}(x))^{n+1}\, d\mu(x)$ can be analysed using the representation theory of the compact Lie group $\mathcal{SO}(k)$. Recall the following facts (see Bröcker and tom Dieck (1985)). Given a Lie group G and representation $\rho_i : G \to \mathcal{GL}_{n_i}(\boldsymbol{R})$ $(i = 1, 2)$, one may construct their tensor product $\rho_1 \otimes \rho_2 : G \to \mathcal{G}_{n_1 n_2}(\boldsymbol{R})$. It is a representation with the property that $\mathrm{tr}\,((\rho_1 \otimes \rho_2)(x)) = \mathrm{tr}\,(\rho_1(x)).\ \mathrm{tr}(\rho_2(x))$. Moreover, if G is compact, and $\rho : G \to \mathcal{GL}_n(\boldsymbol{R})$ is a representation and μ is the Haar probability measure on G, then $\int_G \mathrm{tr}(\rho(x))d\mu(x) = \dim\,(\boldsymbol{R}^n)^G$. Here $(\boldsymbol{R}^n)^G$ denotes the linear subspace of $\boldsymbol{R}^n$ of the elements on which $\rho(x)$ acts trivially for all $x \in G$. In particular denoting by ρ the standard representation $\rho : \mathcal{SO}(k) \hookrightarrow \mathcal{GL}_k(\boldsymbol{R})$ we have $\rho^{\otimes n} = \rho \otimes \ldots \otimes \rho$($n$-fold tensor product) and $(\mathrm{tr}(x))^{n+1} = \mathrm{tr}\,\rho^{\otimes(n+1)}(x)$. Therefore $\int_{\mathcal{SO}(k)} (\mathrm{tr}(x))^{n+1}\, d\mu(x) \geqslant 0$, for all n. Of course the integral is trictly positive if n is odd. Therefore $\alpha > 0$.

Using (3.13) this allows us to write down $M_A(S)$, for by invariance of Haar measure

$$E_A(A^t\, S) = \int_{\mathcal{SO}(k)} A^t\, x \exp\left[\frac{2}{\sigma}\, \mathrm{tr}(A^t\, x)\right] d\mu(x) = M, \qquad \ldots \quad (3.25)$$

so $E_A(S) = AM$ and thus

$$M_A(S) = (E_A^{-1}\,(S))^t\,|\,E_A(S)\,| = AM^{-1}\sqrt{M^2} = A, \qquad \ldots \quad (3.26)$$

as claimed.

Finally let $H \in \boldsymbol{H}_k$, so $AH \in T_A\, \mathcal{SO}(k)$. Then we have (see equation (2.6))

$$(\nabla_{AH}\, l_{(\cdot)})\,(x) = \frac{2}{\sigma}\,(x, AH) = \frac{2}{\sigma}\,\,(S(x), AH), \qquad \ldots \quad (3.27)$$

and also

$$(S(x) - M_A(S), M_A(S)H) = (S(x) - A, AH) = (S(x), AH). \quad \ldots \quad (3.28)$$

Apparently the left hand side terms in (3.27) and (3.38) are proportional. Using the Schwarz inequality as in (2.6) it follows that S attains the lower bound.

Acknowledgement. The authors are very grateful to the referee for his careful reading and constructive comments that led to a considerable improvement of an earlier version of this paper. They would also like to thank Gert Heckman for the proof that $\alpha > 0$ in Section 3.3.

References

BOOTHBY, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Ac. Press, New York.

BRÖCKER, T. & TOM DIECK, T. (1985). *Representations of Compact Lie Groups*, Springer, New York.

CHANG, T. (1986). Spherical regression. *Ann. Statist.* **14**, 907-924.

CURTIS, M. L. (1985). *Matrix Groups*, 2nd ed., Springer, New York.

GRUEB, W. et al. (1972). *Connection, Curvatiure and Cohomology*, Ac. Press, New York.

HENDRIKS, H. (1987). A Cramér-Rao type lower bound for estimators with values in a certain manifold. Report Math Inst., Kath. Un. Nijmegen.

MARDIA, K. V. (1972). *Statistics of Directional Data*, Ac. Press, New York.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York.

SPIVAK, M. (1975). *A Comprehensive Introduction to Differential Geometry*, Vol. IV, Publish or Perish, Boston.

——— (1979). *A Comprehensive Introduction to Differential Geometry*, Vol. II, Publish or Perish, Berkeley.

WATSON, G. S. (1983). *Statistics on Spheres*, Wiley, New York.

DEPARTMENT OF MATHEMATICS
KATHOLIEKE UNIVERSITEIT NIJMEGEN
6525 ED NIJMEGEN
THE NETHERLANDS

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
LUBBOCK, TX 79409
U.S.A.

A 3–14